# Exploring the SDSS Dataset with Linked Scatter Plots: I. EMP, CEMP, and CV Stars

Duane F. Carbon, Christopher Henze, and Bron C. Nelson

*NASA Ames Research Center, NASA Advanced Supercomputing Facility, Moffett Field, CA, 94035-1000, USA;*
*Address correspondence to: Duane.F.Carbon@nasa.gov*

## ABSTRACT

We present the results of a search for EMP, CEMP, and cataclysmic variable stars using a new exploration tool based on linked scatter plots (LSPs). Our approach is especially designed to work with very large spectrum data sets such as the SDSS, LAMOST, RAVE, and Gaia data sets and can be applied to stellar, galaxy, and quasar spectra. As a demonstration, we conduct a search for EMP, CEMP, and cataclysmic variable stars in the SDSS DR10 data set. We first created a 3326-dimensional phase space containing nearly 2 billion measures of the strengths of over 1600 spectral features in 569,738 SDSS stars. These measures capture essentially all the stellar atomic and molecular species visible at the resolution of SDSS spectra. We show how LSPs can be used to quickly isolate and examine interesting portions of this phase space. To illustrate, we use LSPs coupled with cuts in selected portions of phase space to extract EMP stars, C-rich EMP stars, and CV stars. We present identifications for 59 previously unrecognized candidate EMP stars and 11 previously unrecognized candidate CEMP stars. We also call attention to 2 candidate He II emission CV stars found by the LSP approach that have not yet been discussed in the literature.

*Subject headings:* methods: data analysis — stars: abundances — stars: fundamental parameters — surveys — techniques: spectroscopic

## 1. Introduction

With the advent of very large datasets of stellar spectra such as the SDSS, LAMOST, RAVE, and Gaia (e.g. Wyse (2016) for an overview) , it is no longer remotely possible for an observer to digest the contents of an entire dataset visually. The number of spectra is so large that a dedicated expert team could take years to accomplish even a cursory examination. The Henry Draper catalogs (see Cannon & Pickering (1918) and subsequent volumes) are a well known and heroic example which summarized the careful visual examination of "only" 225,000 objects. Because modern datasets can be vastly larger and richer, they are largely resistant to the methodical examination of spectra "by eye" as exemplified by the Draper effort. For this reason, computer-aided approaches for extracting useful information from spectrum archives have become a matter of necessity.

Data mining of large stellar spectrum archives is generally done with very specific, predetermined goals in mind. Typical endpoints are spectral classifications or the estimation of parameters like $T_{\rm eff}$, $\log g$, [Fe/H], and [$\alpha$/Fe]. Many approaches have been used to accomplish these goals; a far less than exhaustive list includes distance minimization (Wilhelm et al. 1999; Allende Prieto et al. 2006; Covey et al. 2007; Lee et al. 2008; Boeche et al. 2011; Bijaoui et al. 2012), projection methods (Recio-Blanco et al. 2006), decision trees (Bijaoui et al. 2012), and artificial neural networks (Bailer-Jones 2000; Re Fiorentin et al. 2007; Manteiga et al. 2010). Often, combinations of these are used in a single data processing pipeline (Lee et al. 2008; Kordopatis et al. 2013; Recio-Blanco et al. 2016).

With all the aforementioned approaches, the researcher must first choose specific examples of spectra, either observed or synthetic, which serve to define the characteristics one hopes to extract from the spectra of the archive. This set of preselected spectra, referred to as a "template" or "training" set, is very problem specific and can be time-consuming to produce. However, once the template/training set is defined and tested, the chosen algorithm can be applied in a fully automated way to each spectrum in the archive to deduce the desired classification or stellar parameters. Large scale application of these techniques has led to major advances in our understanding of galactic chemical evolution by providing very large samples of abundance and kinematic data for analyses in aggregate (e.g., Beers & Carollo 2008; Dierickx et al. 2010; Chiappini et al. 2015; Kordopatis et al. 2016). At the same time, these template/training set techniques have provided a rich pool of information from which researchers may select individual objects for more detailed, higher resolution investigations. These include numerous studies of low metallicity stars (examples referenced in Sections 3.1 and 3.2) as well as efforts to select candidates for such diverse objects as white dwarfs with infalling accretion disks (Gänsicke et al. 2008; Wilson et al. 2014), Li-rich field giants (Martell & Shetrone 2013), and metal poor stars with extreme [$\alpha$/Fe] excesses or deficiencies (Li et al. 2014; Xing & Zhao 2014).

In this paper, we describe an alternative approach to extracting spectra with interesting desired characteristics from large archives. We believe it not only has more innate flexibility and nimbleness than methods using template/training sets but also provides quicker feedback on both the success and failings of the user's choices. To illustrate, we show how the combination of linked scatter plots and the NASA hyperwall were used to data mine a large spectrum dataset, the Sloan Digital Sky Survey, Data Release 10 (SDSS DR10) (Ahn et al. 2014). In particular, we focus on three subsets of interesting stellar objects. First are the extremely metal-poor (EMP) stars whose relevance to the early chemical evolution of our galaxy has made them the focus of considerable effort in the recent years (see Frebel & Norris (2015) for a comprehensive overview). Next are a subset of the EMP stars, the carbon-enhanced EMP (CEMP) stars, which may be informative of supernova element building in the earliest periods of galactic chemical enrichment (Beers & Christlieb 2005; Carollo et al. 2012; Beers & Carollo 2012; Beers 2014; Cooke & Madau 2014; Frebel & Norris 2015). Finally, because the first two examples involved using only absorption features in our search, we demonstrate how one can additionally use emission lines to extract and examine objects like the

cataclysmic variable (CV) stars.

Section 1.1 provides a quick introduction to NASA'a hyperwall and to LSPs. The subsections of Section 2 describe in some detail how we populated our working database of nearly 2 billion feature measurements from the information made available by the SDSS. We discuss how we determined stellar continuum levels (Section 2.1) and how we used these continua to create usable feature measures (Sections 2.2 and 2.3). The reader should be aware that all wavelengths mentioned in this paper are vacuum wavelengths indicated in Angstroms (Å). Section 3.1 details how we used our approach to discover 70 new EMP and CEMP candidates (see Table 1 and Table 2). Section 3.2 recounts how one might extract the 11 new CEMP candidates more directly. In Section 3.3 we show how emission lines can be incorporated in the search as well, in this case to discover CV stars. Section 4 describes shortcomings of our current implementation and outlines needed improvements. Our principal results are summarized in Section 5.

### 1.1. The hyperwall and Linked Scatter Plots

The hyperwall at NASA's Advanced Supercomputing Facility, located at NASA Ames Research Center, is a powerful research tool for the exploration of large observational and theoretical datasets. Here we give a brief introduction to the hyperwall; additional details maybe found in Sandstrom et al. (2003) and Henze (1998). The hyperwall hardware is a $23'$ wide by $10'$ high wall of 128 workstation displays arranged in an $(8 \times 16)$ array, accompanied by a separate operator's console with two displays. Each hyperwall display is controlled by a parallelized software suite running on a node of NASA's Pleiades supercomputer. This software is key to the versatility and exploratory power of the hyperwall. The operator may choose from a variety of modes for using the hyperwall, ranging from display of a complex and active single image involving the whole hyperwall on the one hand to using each of the 128 displays to portray multiple aspects of a dataset on the other. In this study, we employ the latter capability using linked scatter plots.

Linked scatter plots are the single most critical element of our approach and a topic that may be unfamiliar to many astronomers. Here we briefly describe the experiential aspects of using LSPs on the hyperwall. Details of software implementation, which are both highly complex and site-specific, will be sidestepped.

To illustrate our application on the hyperwall, we work with a very large, multi-variable dataset which includes measured feature strengths that we derived from stellar spectra of the SDSS DR10. We downloaded $N_s = 569{,}738$ SDSS spectra and, as we describe below in Section 2, we made 3318 specific feature measurements on each and every spectrum in the set. To these feature measurements, we add seven colors computed from the SDSS *u,g,r,i,z* magnitudes and an estimate of each spectrum's overall signal/noise to make a total of $N_m = 3326$ measured quantities for each spectrum. Each of the $N_m$ quantities is a variable in our multi-variable dataset; each variable contains $N_s$ values of a specific measurement. To appreciate the power of LSPs on the hyperwall, it is useful

to consider the ensemble of data as a phase space with $N_m$ dimensions.

We can populate the hyperwall displays however we wish with $2 \leq N_{hw} \leq 128$ different two-dimensional scatter plots, each plot having a unique coordinate pair drawn from the $N_m$ measurements: ($x_i$, $y_j$; $i$ = 1,2,3,...,$N_m$; $j$ = 1,2,3,...,$N_m$; $i \neq j$). The plots need not share coordinate variables. Armed with the hyperwall's large complement of software tools, the user may choose individual plot scales and centering, view object distributions projected on $x-$ and $y-$axes, and select subsets of plotted points on any display. Subset selection is accomplished by graphical "paint brushes", or by rectangles and lines entered either numerically or by user-manipulated graphical rectangles and triangles.

Because the variables in each of the 128 displays are "linked" by the hyperwall software (Henze 1998), additional powerful, user-guided interactions are possible. For example, if we select a subset of points on any one of the scatter plots (and hence the subset of the objects represented by those points), the points corresponding to the selected objects are automatically highlighted in a distinctive color on all the other scatter plots. This is possible because the points in each scatter plot are "linked" with their corresponding data objects and the hyperwall software tracks the representation of those data objects on all the other displayed scatter plots. The points on each scatter plot are automatically redrawn when a selection occurs. In each instance, the selected points are drawn in the appropriate color, and "on top" of the unselected points to prevent them from being obscured. The hyperwall system is fast enough that drawing the updates is essentially instantaneous, giving the user immediate interactive feedback on the effects of a selection.

The distinctive color aspect of the LSP by itself is very powerful. It allows one to see exactly how the selected points, which may represent a distinctive feature in one scatter plot with a particular pair of variables, are distributed in other scatter plots involving different variables. This extremely useful capability was also used in another graphics tool for astrophysics, VIEWPOINTS, which was adapted from an earlier version of the hyperwall software and developed for single-display platforms (Gazis et al. 2010). Our implementation of the *selection* $\Longleftrightarrow$ *distinctive color* linkage extends for an arbitrary number of successive selections. Thus, having selected a subset of objects in one of the $N_{hw}$ displays, one may then select an additional subset of objects in any of the other ($N_{hw}$ - 1) displays. The newly selected objects will then be rendered in a second distinctive color in all of the displays. *This process may be repeated as often as desired, building up ever more complex selections.*

In a very powerful addition to the multiple selection capability, the hyperwall software allows us to treat the selections as logical Boolean "AND" operations. Thus, for example, we can make a series of selections in different portions of measurement phase space and then show on all the displays with a unique color those objects which simultaneously satisfy all the selection criteria. The LSP software provides full checkpoint and restart capabilities so that the user can safely save and then later restart work involving a series of complex selections on many displays without loss of effort.

At any time in the successive selection process, the user can view on a separate large display at the control console the spectra of the all objects selected up to that time. The spectrum display window gives the user full control over the scales of the spectra, marks the wavelengths of features used in the selections, and allows the user to scroll rapidly through the spectra, examining and intercomparing them. This capability is enormously valuable because it gives the user immediate feedback on how to refine the selection process, and the spectral display window allows the user to quickly identify spectra of particular interest which may be saved for future reference by a simple mouse click.

Having now given an overview of the hyperwall and LSPs, we show in the next section how we selected the $N_s$ spectra and collected the $N_m$ measurements on each spectrum to create our database. This database, in turn, enables us to explore the phase space of feature strengths.

## 2. POPULATING THE HYPERWALL DATABASE

For the purposes of this paper we use optical spectra drawn from Data Release 10 of the Sloan Digital Sky Survey. This choice is for illustrative purposes only. Our approach is generally applicable to any large rich spectrum dataset whether of stars, galaxies, or QSOs. We downloaded as individual FITS files all "SpecObj" spectra of DR10 Class = STAR and zWarning = 0 whose median signal/noise (snMedian) was $\geq 4$. This produced a spectrum dataset containing reasonable quality spectra for 569,738 unique stellar targets. For each of the targets, we also also downloaded such associated quantities as the target's ra, dec, z, zErr, snMedian, and *u,g,r,i,z* psfMag magnitudes. The DR10 data contains spectra from both the SDSS and the BOSS spectrographs, the latter covering a somewhat wider spectral range. The spectrum data includes the vacuum wavelength, calibrated flux ($10^{-17}$ erg cm$^{-2}$ s$^{-1}$ Å$^{-1}$), and pixel-by-pixel inverse-variance values for each object. In order to derive the final dataset that we use on the hyperwall, each spectrum must be processed through a pipeline coded using MATLAB© (MATLAB 2011) whose principal steps we describe next. Before the first pipeline step, all spectrum wavelengths were corrected to zero rest velocity using the value of redshift, z.

### 2.1. Pipeline I — Continuum Determination

In order to make the spectral feature measurements used for the linked scatter plots on the hyperwall, we first must determine a practical reference level, a "continuum", in each spectrum against which measurements are made. We note that this continuum is not the true bound-free/free-free continuum of a stellar atmosphere. Rather, it is an operationally-chosen smooth practical background whose level, for hotter or very metal poor stars, might approach the true continuum but which generally lies well below it. The degree of deviation of the observed continuum from the true continuum depends on the presence of atomic and molecular lines and the spectral resolution.

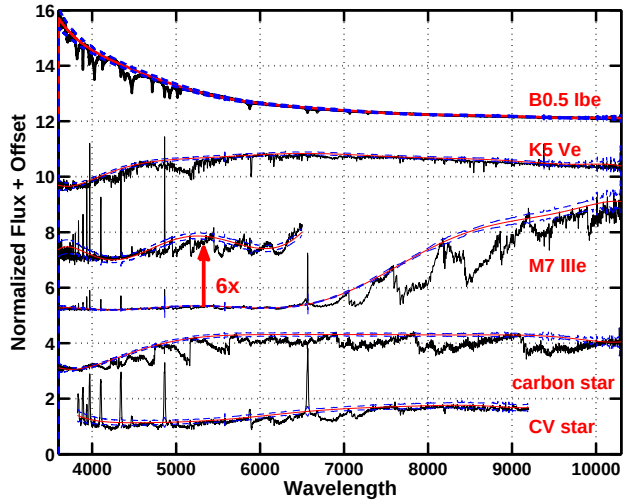


Fig. 1.— Sample automated continuum fitting results with spectral types (SDSS "subClass") noted in red. The automatically fit continua are shown in red; the blue dashed lines show +/− 2.5 $\sigma$ about the continuum computed from the pixel-by-pixel variance. For the sake of plot clarity, the occasional very large values of variance in the M7 IIIe spectrum have been suppressed. The low signal short wavelength end of the M7 IIIe star is offset replotted at 6x for easier examination. Each spectrum is normalized by its average continuum flux and then shifted to avoid overlap.

We need an automated way of placing continua deemed acceptable to a human observer over the very large range of spectral types and abundance variations present in the SDSS DR10 dataset. Moreover, the procedure must be capable of handling the breaks in spectral data which are a not uncommon feature of many DR10 spectra. Our approach largely satisfies these requirements.

We give here an overview of the procedure we employed to determine the continuum for each spectrum in our downloaded DR10 dataset. A detailed description will be given elsewhere and the code provided there as open source. In our procedure we first break the spectrum into intervals of roughly 200 Å width, a number which was chosen empirically so as to work best with the broad molecular absorption features of DR10 cool star spectra. Within each of these intervals, we apply an iterative algorithm that effectively eliminates those wavelengths showing emission significantly above the local noise level. The remaining wavelengths with the highest fluxes are then used to define the continuum level for the interval. Once all intervals are processed, the collection of interval continuum points are fitted with a ninth degree polynomial. If any of the fitting points fall more than two flux standard deviations below the polynomial fit, those fitting points are dropped and the nineth degree polynomial fit is redone with the remaining points. This last step is very effective in preventing the continuum from following large molecular features at the SDSS/BOSS resolutions. The final fit is our adopted continuum for each spectrum.

In Figure 1 we show some typical examples of our automated continuum fits. For clarity the selected objects all have high signal/noise spectra and they represent a variety of objects

posing different challenges to the continuum fitting. A quick examination shows that the deduced continuum nicely follows the overall contours of the spectrum while accurately preserving the innate character of emission and absorption features. This is true even when the spectrum contains mixed atomic emission and broad molecular absorption features. A closer look, however, reveals some disappointing flaws. For example, because we have used a polynomial fit for the final continuum, the continuum level can go awry at the very ends of the spectra, particularly if those spectral regions are comparatively noisy. The continua at the very blue end of the M7 IIIe and CV star spectra are examples of this behavior; in both cases, the continuum is placed higher than one would want for a best fit. Moreover, a close inspection of the detailed fits in Figure 1 reveals interior sections where one might have drawn the continuum differently if working "by eye". However, since such an undertaking for the whole DR10 dataset is quite impossible, we have used our procedure to automatically determine the required 569,738 independent continua. We give an example in Section 3.3 of the type of problem that may created if the continuum fitting is not done well.

### 2.2. Pipeline II — Selecting the Spectral Features to Measure

In the next Section, 2.3, we describe how we measured the spectral feature strengths that make up our hyperwall database. Here we describe how we selected which features to measure in the first place. Since all the objects in our data set are (or at least believed to be) stellar, we aimed to include all principal spectral features (atomic, ionic, and molecular) which might conceivably be measurable at the SDSS/BOSS resolution in some DR10 stars. For the atoms and first ions of all elements which are even moderately abundant in the sun, we downloaded all the parameters necessary to compute individual line absorption strengths from the NIST online tabulation (Kramida et al. 2014). For each line, we computed the value of

$$\lambda^2 \, g_u \, A_{u,l} \, e^{(-1.44*\omega_l/T)} \tag{1}$$

which is proportional to the transition's absorption cross-section, where $\lambda$ is the transition wavelength, $g_u$ is the upper state statistical weight, $A_{u,l}$ is the Einstein A value, $\omega_l$ is the excitation of the transition's lower state in wavenumbers, and $T$ is the temperature in kelvin. The calculation was done at two temperatures: 8200 K and 2895 K; the first is roughly appropriate for the line-forming region in the photosphere of a $T_{\rm eff} = 10000\ K$ star and the latter for the line-forming region of a $T_{\rm eff} = 3640\ K$ star. These $T_{\rm eff}$ correspond to an A0 star (Theodossiou & Danezis 1991) and an M3 III star (Perrin et al. 1998), respectively. For the most abundant elemental species, we selected for inclusion in our feature set all atomic and ionic NIST transitions within the BOSS wavelength range which were at least 1/100 as strong as the strongest transitions at either of the two selected temperatures. For less abundant species, we chose only the strongest lines. To this NIST-based feature set, we added all the atomic and ionic emission lines reported in low resolution spectrum studies by Merrill (1947), Sanford (1949), and Cohen (1980) as well as selected additional lines from Gray & Corbally (2009). We added all the principal band head wavelengths culled from the literature for the the main isotopologues of following molecules: AlO, ScO, TcO, VO, YO, ZrO, AlH,

CaH, CrH, FeH, MgH, SiH, TiH, $C_3$, $SiC_2$, and TiS. Finally, we added the principal band heads from the following isotopologues: $^{12}C^{12}C$, $^{12}C^{13}C$, $^{13}C^{13}C$, $^{12}CH$, $^{13}CH$, $^{12}C^{14}N$, $^{13}C^{14}N$, $^{12}C^{15}N$, $^{47}Ti^{16}O$, $^{48}Ti^{16}O$, $^{49}Ti^{16}O$, and $^{50}Ti^{16}O$. The final feature list contained 1659 wavelengths, $\lambda_i$: 926 atomic and ionic and 733 molecular.

### 2.3. Pipeline III — Measuring the Features

We adopted the feature strength, $S(\lambda_i)$, as our measure of the amount of absorption or emission at wavelength $\lambda_i$:

$$S(\lambda_i) = \frac{[F_c(\lambda_i) - F_l(\lambda_i)]}{F_c(\lambda_i)} \tag{2}$$

where $F_c(\lambda_i)$ is the continuum flux and $F_l(\lambda_i)$ is the observed spectrum's flux at feature wavelength $\lambda_i$. Note that for a fully absorbed spectral line $S(\lambda_i) \rightarrow 1.0$; emission lines may have arbitrarily large negative values of $S(\lambda_i)$. We have also adopted a second feature strength measure, $D(\lambda_i)$, which we have often found helpful when searching for solid detections of weak features:

$$D(\lambda_i) = \left| \frac{[F_c(\lambda_i) - F_l(\lambda_i)]}{\sigma(\lambda_i)} \right| \tag{3}$$

where $\sigma(\lambda_i)$ is an estimate of the noise level at $\lambda_i$ determined from the interpolated pixel-by-pixel inverse-variance for each spectrum. $D(\lambda_i)$ helps to cull out those objects for which the apparent feature is likely to be just noise. $S(\lambda_i)$ and $D(\lambda_i)$ were measured at each of the 1659 feature wavelengths for each of the 569,738 spectra, generating nearly 2 billion measurements from our SDSS DR10 dataset. For clarity in the following text, we will refer to $S(\lambda_i)$ and $D(\lambda_i)$ by a feature's identification rather than its wavelength, e.g., $S(H_\alpha)$ or $S$(Ca II K). When there may be ambiguity, we will indicate the wavelength, e.g., $S$(CH 4308).

In closing this section we want to stress to the reader the simple point that carrying out feature measurements at specific wavelengths does *not* mean one is actually measuring the desired species in all the SDSS stellar spectra. Because the SDSS database contains within itself representatives of nearly all stellar types from hot to cold, oxygen-rich to carbon-rich, considerable confusion can be expected especially given the limited spectral resolution. For example, a measure for He I line absorption at 3889.75 Å, appropriate for hot stars, will be dominated in cooler stars by the blue-degraded $^{12}$CH B-X (0,0) absorption at 3890.10 Å. There are numerous opportunities for such ambiguity and appropriate care must be taken in interpreting phase space selections like those that we describe in the next section.

## 3. APPLICATIONS OF THE HYPERWALL AND LSPs

In the following subsections, we will discuss three straightforward applications of the tools we have described in an effort to make the hyperwall and LSP capabilities more concrete for potential

users. In the first application, we show how LSPs can be used to isolate objects in the DR10 dataset which are candidate EMP and CEMP stars. Next, we will show how one can directly locate candidate CEMP stars. Finally, we will show how one can isolate DR10 objects which are candidate CV stars with He II emission. In each case we make simple selections in some small set of phase spaces to isolate candidate objects of interest. *We emphasize that the selections we make in the following sections are meant only to illustrate the LSP approach. The parameters of our choices are not intended to be definitive.* Experts deeply familiar with the spectra of the objects we seek could easily be expected to make different choices as well as add cuts in other 2-D phase spaces beyond the small set we use. We welcome input from interested researchers.

## 3.1. Application I — EMP Stars

### *3.1.1. Selecting by S/N and u-z*

We begin by illustrating how one might use LSPs to identify EMP stars[1]. While many approaches to this problem based on feature strengths might be devised, the one we use here is deliberately simple. To get an idea of the phase space parameters that might isolate EMP stars, we examined the colors and line strengths for a sample of 168 known SDSS EMP stars discussed in the literature (Aoki et al. 2008; Lai et al. 2009; Behara et al. 2010; Bonifacio et al. 2011; Caffau et al. 2011a,b; Bonifacio et al. 2012; Caffau et al. 2012; Carollo et al. 2012; Sbordone et al. 2012; Aoki et al. 2013; Caffau et al. 2013; Lee et al. 2013; Caffau et al. 2014; Bonifacio et al. 2015; Frebel et al. 2015; Placco et al. 2015).

This review led to the following cuts in feature measurement phase space. Our first cuts were made in the ($u-z$, median S/N) plane, see Figure 2. We chose stars whose median S/N was greater than 35 since it becomes increasingly difficult to make clear judgements as to metallicity when the median S/N falls below that value; the cut in $u$-$z$ color is guided by the $u$-$z$ range of the 168 known SDSS EMP stars. This first cut selects roughly 10% of the stars in the whole dataset.

### *3.1.2. Subselecting by $S(\mathrm{H}_\epsilon)$ and $S$(Ca II K)*

Next, we turn our attention to the ($S(\mathrm{H}_\epsilon)$, $S$(Ca II K)) plane. It is instructive to first examine this portion of feature strength phase space with the stars selected using the ($u$-$z$, median S/N) cut superimposed over all the stars in the relevant portion of the ($S(\mathrm{H}_\epsilon)$, $S$(Ca II K)) plane, Figure 3.

In addition to the general placement of our selected subset in the ($S(\mathrm{H}_\epsilon)$, $S$(Ca II K)) plane, two particular points should be noted. First, given the values of N in this figure and in Figure 2, not all stars selected by the ($u$-$z$, median S/N) cut lie within the bounds of the current figure; if

[1]We adopt the definition of Beers & Christlieb (2005, Table I)

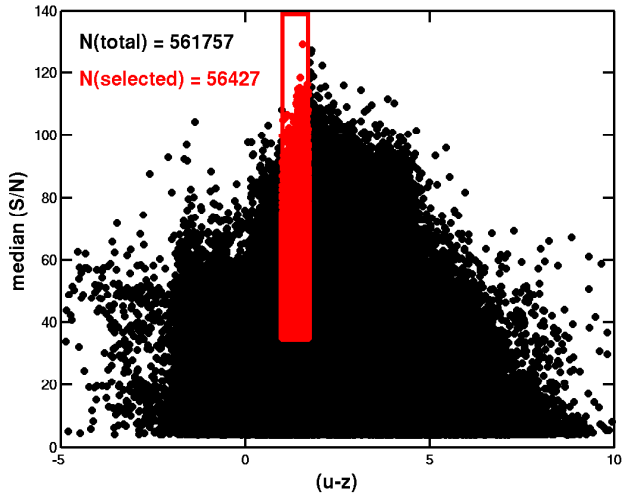


Fig. 2.— Culling by ($u$-$z$) and S/N. All the downloaded stars within the figure boundaries are shown in black. The selection (in red) extracts the subset of those stars that satisfy the constraints ($1.0 \leq (u\text{-}z) \leq 1.7$) and (median S/N $> 35$). The color-coded "N"s in the figure indicate how many stars are in the figure area and how many are sub-selected at this stage.

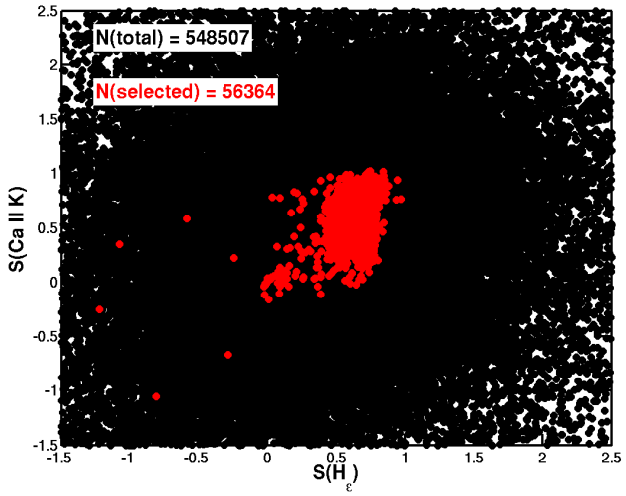


Fig. 3.— Distribution in the ($S(\mathrm{H}_\epsilon)$, $S$(Ca II K)) plane of stars previously selected in the ($u$-$z$, median S/N) plane. All the downloaded SDSS stars within the figure boundaries are shown in black. Those selected by the cuts (($1.0 \leq (u\text{-}z) \leq 1.7$), (median S/N $> 35$)) in the ($u$-$z$, median S/N) plane are shown in red. The color-coded "N"s indicate how many stars are in the figure area and how many in the area were previously selected.

desired, this could be cured by expanding the plot domain. Second, clearly many of the objects in Figure 3 have $S(\lambda_i) > 1.0$, fully 8.4% of the population seen in this figure. Naively one might have expected such values to be an impossibility given the definition of $S(\lambda_i)$. These aberrant values occur because when the signal level in a SDSS spectrum is small, negative values of flux often occur. These, when combined with positive values for the continuum flux consistently yield $S(\lambda_i) > 1.0$. This effect is particularly exacerbated when $F_c(\lambda_i)$ approaches zero in the denominator of Equation 2. It should be noted that the stars which satisfy the median S/N > 35 constraint do not show these anomalous values since higher median S/N generally means higher signal strength in the blue.

Now, examining the known SDSS EMP stars, we look for parameters which will isolate those stars most likely to be EMP candidates from the general population. Figure 4 compares the SDSS spectrum of a extreme EMP star (J130017.20+263238.6 (6393, 4.00, [M/H] = -3.65) Bonifacio et al. (2012)) with a near solar metallicity star (HD 043042 (6392, 4.22, [M/H] = 0.08) Sánchez-Blázquez et al. (2006)) from the MILES Library. Since the effective temperatures and surface gravities of the two stars agree so closely, differences in appearance are due to metallicity alone. The dramatic weakening of the metal lines in the EMP star is clearly evident. It is immediately obvious that the signature low ratio of $S(\text{Ca II K})/S(\text{H}_\epsilon)$[2] might be a useful selection constraint. For J130017.20+263238.6 this ratio is 0.39, whereas the ratio is $\approx 1.0$ for HD 043042. While many EMP stars have $S(\text{Ca II K})$ comparable to, and even larger, than their $S(\text{H}_\epsilon)$, stars in the ($S(\text{H}_\epsilon)$, $S(\text{Ca II K})$) plane that satisfy the constraint $S(\text{Ca II K}) < 0.6 \times S(\text{H}_\epsilon)$ are very likely to be EMP stars. This cut is shown in Figure 5 where we have restricted ourselves to the relevant absorption portion ($0.0 \leq S(\lambda_i) \leq 1.0$) of the ($S(\text{H}_\epsilon)$, $S(\text{Ca II K})$) plane and show only those stars that already satisfy the ($u$-$z$, median S/N) cut. For reference, we also show in Figure 5 the positions of the known SDSS EMP stars mentioned above. Close examination reveals that only 33% of the known EMP stars lie within the currently selected region. Fully two-thirds of known SDSS EMP stars have $S(\text{Ca II K}) > 0.6 \times S(\text{H}_\epsilon)$; indeed, 23% lie above the line defined by $S(\text{Ca II K}) = S(\text{H}_\epsilon)$. Thus, while those currently selected are excellent candidates, it is clear that different constraints would add to the pool of candidates. We chose not to broaden the search area primarily because with our current constrained variables, $u$-$z$, median S/N, $S(\text{Ca II K})$, and $S(\text{H}_\epsilon)$, it was not clear to us how to disambiguate EMP stars from stars with higher metallicity on the basis of our current measurement set. Perhaps a cleverer set of constraints than the ones we have chosen would be able to net a larger sample of high likelihood EMP candidates. Again we encourage suggestions from our readers.

[2]Note that because of the limited SDSS resolution, the depth of $\text{H}_\epsilon$ is impacted by the presence of the unresolved, overlapping Ca II H line. Nevertheless, $\text{H}_\epsilon$ clearly dominates for the stars we care about.

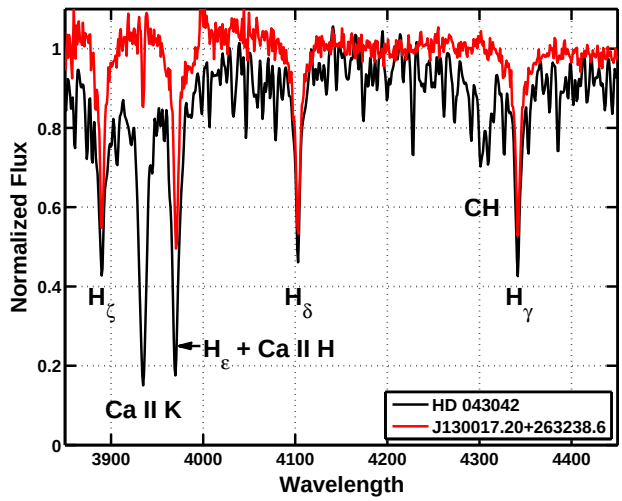


Fig. 4.— Comparison of the SDSS spectrum of a known EMP star (J130017.20+263238.6 in red) and a solar abundance star (HD 043042 in black). The stars have essentially the same $T_{\rm eff}$ and $\log g$. Each spectrum is normalized by its average flux over the interval [4194,4195]. Atomic and molecular features prominent in EMP or CEMP spectra at the SDSS spectral resolution are specifically indicated.

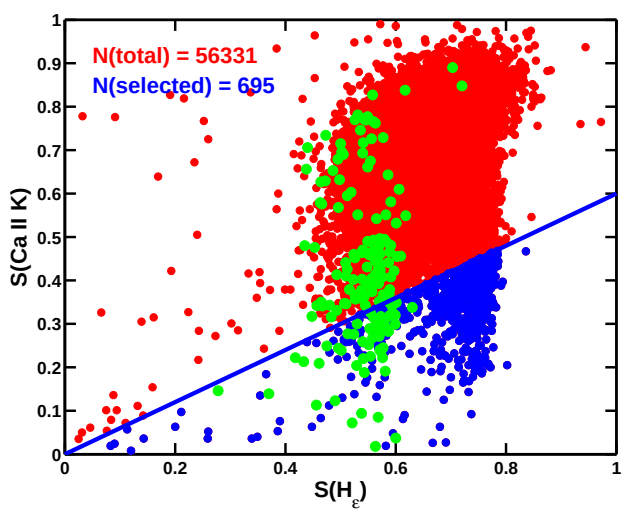


Fig. 5.— Culling potential EMP candidates. All the red and blue stars in the figure satisfy the constraints: $(1.0 \leq (u\text{-}z) \leq 1.7)$ and (median S/N $> 35$). The stars indicated in blue *additionally* satisfy the constraint that $(S(\text{Ca II K}) < 0.6 \times S(\text{H}_\epsilon))$, a restriction that makes the selected stars likely EMP candidates. The color-coded "N"s indicate how many stars are in the figure area and how many were sub-selected by the diagonal cut. For reference the 166 known SDSS EMP stars that have both $S(\text{Ca II K})$ and $S(\text{H}_\epsilon)$ measures are shown in green.

### *3.1.3. Eliminating obvious false positives - step 1*

The described cuts in the ($u$-$z$, median S/N) and ($S(\mathrm{H}_\epsilon)$, $S$(Ca II K)) phase spaces isolated 695 EMP candidates (0.12% of all the stars) whose spectra were then examined visually in two steps. The first step requires only a brief visual examination of each spectrum and is accomplished quite quickly. This step is crucial in weeding out two groups of stars: stars with badly flawed SDSS spectra at $\mathrm{H}_\epsilon$ and/or Ca II K, and stars with good spectra which matched the selection criteria but which were obviously not classical EMP stars. The latter were either early type stars (late B, early A), distinguished by their broad H line wings and steep SEDs, or white dwarf + cool star binaries/superpositions. After this first culling, 139 stars remained as apparently possible EMP candidates. It is important to note that 30 of these 139 have already been established as EMP stars through model atmosphere abundance analysis; 21 by Aoki et al. (2013), and the rest by Caffau et al. (2011b), Caffau et al. (2011a), Caffau et al. (2012), Bonifacio et al. (2012), Sbordone et al. (2012), Caffau et al. (2013) and Li et al. (2014). *By itself, this is a strong independent confirmation that careful selections even in the limited 4-dimensional phase space we have considered can have real discovery power.* We find it particularly satisfying that one of the known EMP stars captured by our selections is J102915.15+172928.0, currently the most metal-deficient SDSS EMP star known (Caffau et al. 2011a, 2012).

### *3.1.4. Eliminating more subtle false positives - step 2*

Since a detailed model atmosphere abundance analysis of these objects is well outside the scope of this paper, we attempted to decide which of the remaining 109 (corrected for the 30 known EMP stars) stars were the most likely EMP candidates by systematically comparing their SDSS spectral energy distributions (SEDs) to those of 168 known SDSS EMP stars. The set of known SDSS EMP stars is not without shortcomings unfortunately. First, the stars in this set often have SDSS spectra which are quite noisy or have poor correction for terrestrial absorption/emission. These defects often make them poor comparison stars. Second, while classified as EMP stars by the authors of the papers making the abundance analyses, their metallicities do not necessarily adhere to the metallicity class definitions laid down by Beers & Christlieb (2005, Table 1) and now widely adopted. Thus, a significant number of stars in the sample of 168 known SDSS EMP stars are probably more likely on the metal-poorer end of Beers and Christlieb's "very metal poor" classification. However, given the often large differences in metallicity (as well as in $T_{\mathrm{eff}}$ and log $g$) that one encounters for the same star as analyzed by different investigators, it is difficult to be too strict in assigning a star to a particular metallicity classification based on the current literature. As a consequence, we adopted the attitude that if a star has the spectral characteristics, as defined below, of stars in the sample of 168 known SDSS EMP stars, we shall select it as a "likely" EMP candidate star. Those appropriately armed with high resolution spectra and careful model atmosphere abundance analyses will have the final word.

The spectrum of each of the 109 stars in the remaining candidate list was compared with the spectra of known SDSS EMP stars. This somewhat time-consuming task was made considerably more tractable by using the SCAMP spectrum visualization tool developed here (Idehara & Carbon 2004). Particular attention was paid to the fit of the known and candidate SEDs over the entire spectral range. This helped us weed out luminous A stars which have $S(\text{Ca II K})/S(\text{H}_\epsilon)$ ratios in our selected domain but significantly steeper SEDs than the known EMP stars. For objects with closely fitting SEDs, careful attention was then given to how well the depths and widths of the H Balmer lines agreed between the candidate and known EMP stars. We also affirmed that any other metal lines appearing in the spectrum were consistent with the strengths in similar known SDSS EMP stars. If the overall SED and Balmer lines fits were close, other metal lines appropriately weak or absent, and the Ca II K line in the candidate spectrum was close to or weaker than those in the spectra of (generally) two or more matching known SDSS EMP stars, we selected the candidate star as a "likely" EMP candidate. Occasionally, we had to relax these criteria slightly usually because the closest matching known SDSS spectra were too noisy for a clear determination.

#### *3.1.5. The likely EMP and CEMP stars*

Table 1 identifies the 57 candidate EMP stars chosen for their similarity to known EMP stars. Along with their standard SDSS star name, we include the plate, fiber, and modified Julian dates (MJD) of the observed spectrum we used, the $g$ magnitude, the $u$-$g$ and $g$-$r$ colors, relevant $S(\lambda_i)$ and $D(\lambda_i)$ values, the spectral type (SpT) assigned by the SDSS pipeline ("subClass"), plus any additional comments regarding the star's spectrum. The apparent presence of the $^{12}$CH A-X (0,0) band is explicitly noted. It should be observed that the most frequently occurring spectral type in Table 1 is "A0" with 42 stars (74%) so assigned. This is not much different than the result for the 168 known SDSS EMP stars where 66% of the stars were assigned the "A0" spectral type by the SDSS. This suggests that the "A0" stars in the SDSS dataset could be a fertile search area for additional EMP and CEMP stars provided one uses the right search constraints. Since there are 56,661 stars classified as "A0" in our downloaded dataset, the importance of good constraints for EMP/CEMP sub-selection is clear.

Table 1. Candidates Likely to be EMP stars

| Star Name | Plate-Fiber-MJD | $g$ | $u$-$g$ | $g$-$r$ | $S$ Ca II K | $D$ Ca II K | $S$ $H_\epsilon$ | $D$ $H_\epsilon$ | $S$ CH 4308 | $D$ CH 4308 | SpT | Notes |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| J000401.20+154757.6 | 0751-390-52251 | 16.71 | 0.84 | 0.24 | 0.35 | 11.92 | 0.60 | 26.68 | 0.06 | 1.98 | A0 | ··· |
| J002942.59−023723.9 | 4367-62-55566 | 16.70 | 0.84 | 0.25 | 0.28 | 15.21 | 0.53 | 36.30 | 0.05 | 2.92 | A2II | ··· |
| J011844.41+111741.5 | 4667-34-55868 | 16.98 | 0.84 | 0.25 | 0.23 | 8.52 | 0.53 | 25.14 | 0.02 | 0.64 | A1III | ··· |
| J012822.15+225013.4 | 5107-428-55940 | 16.28 | 0.84 | 0.23 | 0.27 | 11.89 | 0.51 | 26.33 | 0.24 | 14.65 | F2III | CH |
| J025052.23−032452.0 | 4341-449-55538 | 16.43 | 0.83 | 0.25 | 0.30 | 14.77 | 0.53 | 32.63 | 0.06 | 3.24 | F0V | ··· |
| J050110.33−033955.1 | 3209-472-54906 | 17.51 | 0.82 | 0.22 | 0.27 | 10.99 | 0.58 | 26.87 | 0.04 | 1.83 | A0 | ··· |
| J061136.45+840554.3 | 2540-509-54110 | 16.31 | 0.79 | 0.25 | 0.29 | 10.49 | 0.56 | 27.39 | 0.05 | 2.16 | A0 | ··· |
| J073935.14+443442.4 | 3663-86-55176 | 16.47 | 0.84 | 0.24 | 0.26 | 14.66 | 0.52 | 37.18 | 0.05 | 3.17 | F0V | ··· |
| J074027.05+303339.3 | 4448-440-55538 | 17.13 | 0.89 | 0.28 | 0.33 | 13.50 | 0.56 | 28.11 | 0.03 | 1.08 | B3II | ··· |
| J074748.62+264543.5 | 2055-121-53729 | 17.12 | 0.92 | 0.20 | 0.17 | 4.99 | 0.55 | 22.07 | 0.03 | 1.15 | A0 | ··· |
| J074943.30+670825.1 | 2939-561-54515 | 16.32 | 0.80 | 0.24 | 0.33 | 17.16 | 0.62 | 39.98 | 0.05 | 3.16 | A0 | ··· |
| J075731.03+120716.7 | 2265-239-53674 | 16.87 | 0.86 | 0.21 | 0.33 | 9.93 | 0.57 | 17.96 | 0.06 | 1.96 | A0 | ··· |
| J080326.35+325146.9 | 3756-688-55505 | 17.23 | 0.89 | 0.29 | 0.29 | 11.20 | 0.53 | 24.86 | 0.08 | 3.34 | A2II | ··· |
| J080336.58+053430.6 | 2056-190-53463 | 16.93 | 0.83 | 0.23 | 0.29 | 8.12 | 0.54 | 18.45 | 0.05 | 1.66 | A0 | ··· |
| J080711.80+151458.9 | 4499-608-55572 | 17.37 | 0.84 | 0.26 | 0.25 | 8.15 | 0.53 | 21.11 | 0.10 | 3.89 | A2II | CH |
| J081554.26+472947.9 | 3693-408-55208 | 16.79 | 0.85 | 0.29 | 0.11 | 3.53 | 0.48 | 19.37 | 0.30 | 14.33 | F3/F5V | CH |
| J081754.08+451035.6 | 3691-30-55274 | 17.55 | 0.85 | 0.24 | 0.29 | 9.25 | 0.52 | 18.30 | 0.16 | 5.90 | A1III | CH |
| J084932.30+583458.3 | 1784-601-54425 | 16.93 | 0.89 | 0.25 | 0.32 | 11.39 | 0.57 | 25.57 | 0.06 | 2.18 | A0 | ··· |
| J094143.13+361028.3 | 3223-170-54865 | 17.04 | 0.89 | 0.20 | 0.33 | 14.11 | 0.61 | 33.81 | 0.05 | 2.33 | A0 | ··· |
| J094147.00−030131.2 | 3782-282-55244 | 16.77 | 0.81 | 0.25 | 0.31 | 14.13 | 0.56 | 33.10 | 0.08 | 4.33 | A2II | ··· |
| J094940.20+270703.6 | 2342-234-53742 | 16.06 | 0.82 | 0.28 | 0.36 | 13.99 | 0.61 | 32.18 | 0.28 | 13.47 | A0 | CH |
| J101217.52+262808.5 | 2386-557-54064 | 16.99 | 0.84 | 0.21 | 0.33 | 13.56 | 0.56 | 29.41 | 0.04 | 1.44 | A0 | ··· |
| J105613.05+451845.1 | 4689-708-55656 | 16.93 | 0.78 | 0.26 | 0.32 | 14.34 | 0.55 | 30.41 | 0.08 | 3.54 | A1III | ··· |
| J105928.93+470710.3 | 2390-2-54094 | 17.53 | 0.88 | 0.21 | 0.13 | 3.25 | 0.56 | 19.08 | 0.06 | 1.93 | A0 | ··· |
| J110821.20−164648.2 | 2690-165-54211 | 16.38 | 0.90 | 0.21 | 0.36 | 10.11 | 0.62 | 17.31 | 0.08 | 2.70 | A0 | ··· |
| J112147.64−120842.0 | 2859-106-54570 | 16.19 | 0.85 | 0.18 | 0.24 | 10.06 | 0.59 | 27.99 | 0.19 | 10.48 | A0 | CH |
| J113453.71+663943.5 | 2858-111-54498 | 17.69 | 0.93 | 0.18 | 0.24 | 6.71 | 0.57 | 19.94 | 0.01 | 0.34 | A0 | [a] |
| J115315.67+313056.0 | 1991-167-53446 | 16.67 | 0.81 | 0.24 | 0.18 | 5.18 | 0.55 | 20.15 | 0.08 | 2.78 | A0 | ··· |
| J120253.60−002620.4 | 2892-194-54552 | 15.77 | 0.81 | 0.20 | 0.26 | 13.61 | 0.57 | 38.13 | 0.03 | 1.36 | A0 | ··· |
| J121619.51+494126.3 | 2919-31-54537 | 17.94 | 0.83 | 0.24 | 0.35 | 7.46 | 0.64 | 16.39 | 0.05 | 1.46 | A0 | [b] |
| J124643.01+484453.1 | 2898-288-54567 | 16.89 | 0.79 | 0.26 | 0.33 | 13.45 | 0.56 | 26.14 | 0.05 | 2.49 | A0 | ··· |
| J124947.16+502202.2 | 2898-468-54567 | 17.33 | 0.84 | 0.20 | 0.36 | 11.56 | 0.61 | 24.99 | 0.07 | 2.58 | A0 | [b] |
| J125005.10+102156.4 | 2963-474-54589 | 16.29 | 0.94 | 0.26 | 0.33 | 15.44 | 0.54 | 34.05 | 0.02 | 0.94 | A0 | ··· |

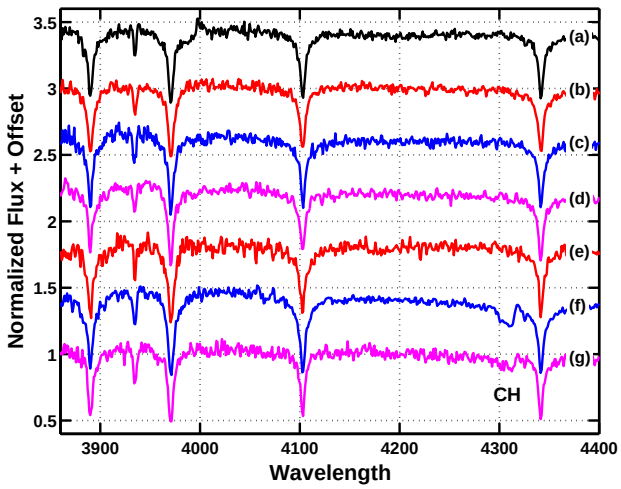


Fig. 6.— Comparison of a known SDSS EMP with candidate EMP/CEMP stars from Table 1 which have similar SEDs. The spectra have been normalized to their mean flux in the interval [4150,4270]. The stars shown are: (a) J130017.20+263238.6, a known SDSS EMP star; (b) J011844.41+111741.5; (c) J074748.62+264543.5; (d) J115315.67+313056.0; (e) J160257.13+044628.8; (f) J112147.64-120842.0; (g) J080711.80+151458.9. The last two stars have discernible $^{12}$CH A-X (0,0) bands.

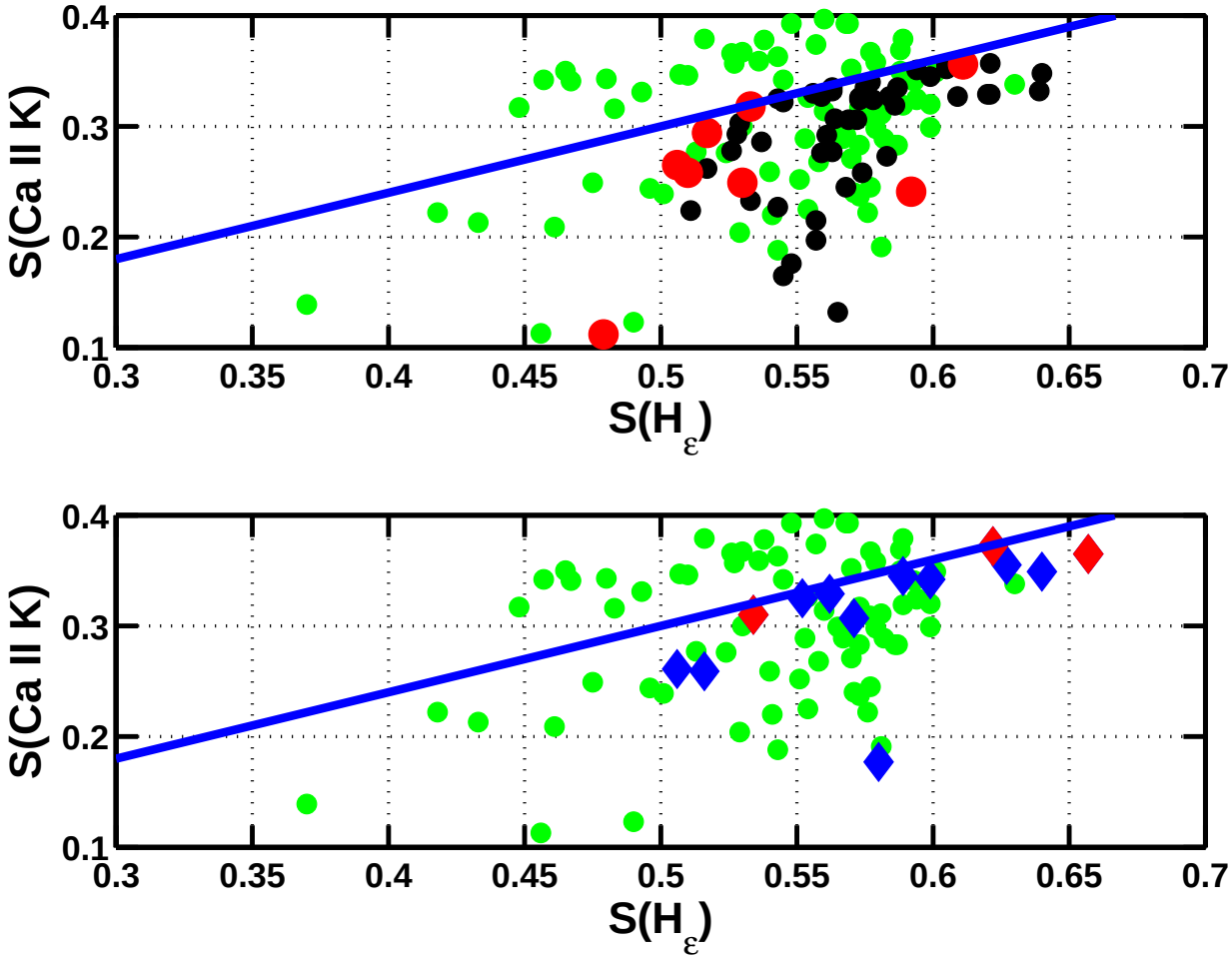


Fig. 7.— Distribution of likely and uncertain candidates. Note that the scales differ from those in Figure 5. In the upper plot the stars represented by black dots are the EMP candidates from Table 1 while the stars represented by red dots are the CEMP candidates in that Table. Similarly in the lower plot, the stars represented by blue diamonds are the EMP candidates from Table 2 while the stars represented by red diamonds are the CEMP candidates in that Table. For reference the 166 known SDSS EMP stars that have both $S$(Ca II K) and $S(\mathrm{H}_\epsilon)$ measures are shown in green in both the upper and lower plots.

We show in Figure 6 some examples from Table 1 that were chosen because their SEDs were similar to the SED of a well-established EMP, J130017.20+263238.6. Two independent studies of J130017.20+263238.6 found very similar stellar parameters ($T_{\rm eff}$,$\log g$,[Fe/H]): (6393, 4.00, -3.65) from Bonifacio et al. (2012) and (6450, 4.0, -3.53) from Aoki et al. (2013). Two of these stars, J080711.80+151458.9 and J112147.64-120842.0, show evidence of $^{12}$CH A-X (0,0) absorption and therefore are candidate CEMP stars.

### *3.1.6. Uncertain EMP candidates and rejected stars*

Not all of the 109 stars selected for closer examination fell nicely within the boundaries of spectral characteristics defined by the 168 known SDSS EMP stars. Thirteen stars were found to be arguably close to the spectrum space defined by the 168 known SDSS EMP stars but could not convincingly be shown to lie comfortably within that group. The problem often lay in the fact that the spectrum of the candidate EMP and/or the spectra of the closest known EMP stars were too noisy to make a clean judgment. In a number of cases, a single 100 - 300 Å stretch of a particular candidate's SED did not fit by a wide margin the SEDs of any potentially matching known EMP stars. We suspect localized poor fluxing by the SDSS pipeline in these cases. There were also a few stars which had poor SED fits or poor Balmer line fits with known EMP stars but showed weak Ca II K absorption and CH absorption. We classified all 13 stars as "uncertain" and show their basic data in Table 2. These stars warrant some consideration since they may be EMP stars, or even CEMP stars. However, we feel the likelihood is not as high as for those in Table 1.

Before closing this subsection, we wish to note the unexpected SDSS classification of J202109.02+601605.3 as a cataclysmic variable. While there are stars in the Szkody tabulation (see Section 3.3) of CV stars that do have SEDs somewhat like that of J202109.02+601605.3, they all show prominent emission lines. The spectrum of our candidate EMP star appears to have no emission lines at all. Disregarding its emission lines, the CV J204448.92-045928.8 is the closest match in the overall SED shape. However, since its secondary is estimated to be be K4 or K5 (Peters & Thorstensen 2005), it displays numerous strong absorption features throughout its SDSS spectrum not apparent in the much more featureless J202109.02+601605.3. On the other hand, J202109.02+601605.3's SED and Balmer line widths/depths are bracketed by two known SDSS EMP stars, J230959.55+230803.1 and J170339.60+283649.9. For this reason, we offer J202109.02+601605.3 as an "uncertain" EMP candidate.

### *3.1.7. Distribution of likely and uncertain candidates*

We close with a brief examination of the distribution of the stars in Tables 1 and 2 in the ($S$(H$_\epsilon$),$S$(Ca II K)) plane. Figure 7 shows how the stars chosen for inclusion in those Tables are distributed relative to all the stars selected by the diagonal cut in Figure 5. Perhaps the most

Table 1—Continued

| Star Name | Plate-Fiber-MJD | $g$ | $u$-$g$ | $g$-$r$ | $S$ Ca II K | $D$ Ca II K | $S$ $H_\epsilon$ | $D$ $H_\epsilon$ | $S$ CH 4308 | $D$ CH 4308 | SpT | Notes |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| J125203.88+500015.3 | 2898-498-54567 | 16.42 | 0.88 | 0.21 | 0.32 | 19.01 | 0.59 | 47.06 | 0.06 | 3.55 | A0 | ··· |
| J125246.46+205516.4 | 5988-46-56072 | 17.24 | 0.77 | 0.25 | 0.22 | 7.01 | 0.51 | 19.99 | 0.04 | 1.67 | A1III | ··· |
| J125556.07+101715.1 | 3234-355-54885 | 17.67 | 0.84 | 0.22 | 0.32 | 8.41 | 0.53 | 16.21 | 0.25 | 8.85 | A0 | CH |
| J130149.62+360023.7 | 2016-350-53799 | 16.36 | 0.89 | 0.20 | 0.33 | 11.68 | 0.57 | 26.87 | 0.07 | 2.73 | A0 | ··· |
| J130701.00+345611.2 | 2016-247-53799 | 16.59 | 0.84 | 0.22 | 0.34 | 10.79 | 0.60 | 21.82 | 0.07 | 2.27 | A0 | ··· |
| J131116.58+001237.6 | 2901-464-54652 | 16.20 | 0.90 | 0.14 | 0.20 | 6.69 | 0.56 | 27.91 | 0.04 | 1.60 | A0 | [c] |
| J132127.81+173341.5 | 2605-474-54484 | 16.35 | 0.85 | 0.19 | 0.35 | 16.94 | 0.59 | 39.16 | 0.05 | 2.23 | A0 | ··· |
| J133224.76+273058.2 | 2245-450-54208 | 16.78 | 0.83 | 0.22 | 0.23 | 7.34 | 0.54 | 23.73 | 0.06 | 2.06 | A0 | ··· |
| J134003.05+095755.8 | 2928-412-54614 | 16.59 | 0.85 | 0.19 | 0.32 | 22.13 | 0.58 | 51.28 | 0.05 | 3.13 | A0 | ··· |
| J134648.05+273145.4 | 2904-136-54574 | 15.97 | 0.86 | 0.20 | 0.31 | 27.14 | 0.57 | 60.19 | 0.05 | 4.17 | A0 | ··· |
| J134847.62+410805.0 | 1377-571-53050 | 16.78 | 0.85 | 0.22 | 0.35 | 11.97 | 0.60 | 25.30 | 0.04 | 1.23 | A0 | ··· |
| J134913.28+173602.7 | 2905-133-54580 | 16.83 | 0.81 | 0.18 | 0.34 | 12.80 | 0.59 | 23.87 | 0.05 | 2.06 | A0 | ··· |
| J143303.99+205157.2 | 2964-306-54632 | 16.52 | 0.85 | 0.18 | 0.33 | 12.78 | 0.57 | 25.63 | 0.05 | 2.04 | A0 | ··· |
| J143943.92−021051.1 | 0919-78-52409 | 16.50 | 0.85 | 0.21 | 0.33 | 10.13 | 0.64 | 26.06 | 0.01 | 0.45 | A0 | ··· |
| J144811.72+384601.9 | 1350-67-52786 | 16.64 | 0.85 | 0.21 | 0.33 | 10.44 | 0.62 | 26.06 | 0.08 | 2.86 | A0 | ··· |
| J145409.78−001111.0 | 3314-206-54970 | 17.31 | 0.87 | 0.18 | 0.26 | 7.31 | 0.51 | 17.89 | 0.20 | 7.77 | A0 | CH |
| J151932.49+063610.7 | 2902-104-54629 | 16.48 | 0.85 | 0.17 | 0.34 | 17.42 | 0.58 | 35.08 | 0.04 | 2.13 | A0 | ··· |
| J154045.84+034914.7 | 0594-349-52045 | 17.00 | 0.84 | 0.25 | 0.33 | 9.72 | 0.58 | 21.58 | 0.09 | 3.29 | A0 | [b] |
| J154811.21+261450.9 | 2459-309-54544 | 17.05 | 0.84 | 0.20 | 0.31 | 12.59 | 0.57 | 25.19 | 0.04 | 1.71 | A0 | ··· |
| J160257.13+044628.8 | 2175-294-54612 | 17.27 | 0.89 | 0.20 | 0.28 | 7.32 | 0.56 | 16.89 | 0.08 | 2.66 | A0 | [b] |
| J160443.30+462513.6 | 0813-140-52354 | 16.55 | 0.84 | 0.22 | 0.34 | 11.55 | 0.57 | 21.50 | 0.04 | 1.52 | A0 | ··· |
| J173239.50+640049.1 | 2551-164-54552 | 16.71 | 0.83 | 0.20 | 0.21 | 6.46 | 0.56 | 21.57 | 0.05 | 1.72 | A0 | ··· |
| J220055.92+095253.5 | 5065-476-55739 | 17.87 | 0.79 | 0.26 | 0.28 | 9.74 | 0.56 | 24.41 | 0.05 | 1.88 | A2II | ··· |
| J231031.86+031847.6 | 4288-862-55501 | 16.84 | 0.80 | 0.25 | 0.34 | 11.20 | 0.56 | 23.95 | 0.01 | 0.22 | A2II | [b] |

[a]Steepest SED slope of all candidate EMP stars, possible comparison spectra are quite noisy.

[b]Candidate has noisy SDSS spectrum.

[c]Very weak Ca II K line.

Table 2. Stars with "Uncertain" Classification

| Star Name | Plate-Fiber-MJD | $g$ | $u$-$g$ | $g$-$r$ | $S$ Ca II K | $D$ Ca II K | $S$ $H_\epsilon$ | $D$ $H_\epsilon$ | $S$ CH 4308 | $D$ CH 4308 | SpT | Notes |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| J112736.01+592431.3 | 3211-235-54852 | 17.57 | 0.89 | 0.15 | 0.35 | 10.26 | 0.63 | 23.81 | 0.11 | 4.28 | A0 | ··· |
| J121925.37−000848.6 | 2558-315-54140 | 16.98 | 0.87 | 0.19 | 0.37 | 15.01 | 0.62 | 30.30 | 0.19 | 8.96 | A0 | CH |
| J130619.63+394942.4 | 2900-390-54569 | 16.55 | 0.93 | 0.09 | 0.36 | 13.08 | 0.66 | 31.14 | 0.16 | 6.41 | A0 | CH |
| J133936.19+103304.7 | 2903-404-54581 | 14.90 | 0.85 | 0.20 | 0.34 | 17.89 | 0.60 | 46.47 | 0.05 | 2.34 | A0 | ··· |
| J141022.25+430151.5 | 6058-980-56090 | 16.78 | 0.84 | 0.26 | 0.31 | 11.88 | 0.53 | 25.30 | 0.15 | 5.83 | A2II | CH |
| J144501.72−010925.4 | 2909-282-54653 | 15.95 | 0.87 | 0.18 | 0.31 | 22.02 | 0.57 | 48.43 | 0.04 | 2.60 | A0 | ··· |
| J152706.80+385509.0 | 2911-146-54631 | 17.24 | 0.87 | 0.21 | 0.34 | 15.81 | 0.59 | 24.88 | 0.08 | 3.82 | A0 | ··· |
| J152735.78+383933.4 | 2936-97-54626 | 17.91 | 0.87 | 0.18 | 0.18 | 4.59 | 0.58 | 18.08 | 0.08 | 2.77 | A0p | [a] |
| J155159.34+253900.6 | 2459-256-54544 | 17.11 | 0.91 | 0.27 | 0.26 | 10.30 | 0.52 | 19.85 | 0.07 | 2.97 | A0 | ··· |
| J173632.66+643218.9 | 2561-559-54597 | 18.08 | 0.86 | 0.23 | 0.35 | 9.41 | 0.64 | 21.66 | 0.07 | 2.40 | A0 | ··· |
| J202109.02+601605.3 | 2564-596-54275 | 16.98 | 0.75 | 0.34 | 0.33 | 8.94 | 0.55 | 18.93 | 0.05 | 1.84 | CV | ··· |
| J220012.99+092501.4 | 4095-8-55497 | 17.48 | 0.80 | 0.23 | 0.26 | 8.86 | 0.51 | 20.48 | 0.10 | 3.86 | A1III | ··· |
| J235252.78+055513.4 | 4405-688-55854 | 17.43 | 0.83 | 0.27 | 0.33 | 10.52 | 0.56 | 22.86 | 0.08 | 2.93 | A2II | ··· |

[a]Markedly weak Ca II K line.

striking aspect of the distribution of the chosen candidates is that none of them have $S(\mathrm{H}_\epsilon) > 0.66$ whereas the distribution of the selected (blue) stars in Figure 5 extends to $S(\mathrm{H}_\epsilon) = 0.84$. In fact, the region with $S(\mathrm{H}_\epsilon) > 0.66$ contains most of the selected stars, fully 500 of the 695 selected, or 72 %. The majority of the stars in this region were deleted because they were white dwarfs, white dwarf/cool star binaries, or early type stars with very broad H Balmer lines. Larger values of $S(\mathrm{H}_\epsilon)$ generally go hand-in-hand with increasing line breadth. The rest were deleted because they had defective spectra. Most were deleted in the coarse cull described in Section 3.1.3, the rest (18) were deleted in the fine cull of Section 3.1.4.

The EMP and CEMP candidates from Table 1 shown in the upper plot of Figure 7 appear to follow the roughly same distribution as the known EMP/CEMP stars. The situation seems somewhat different in the lower plot where we see the distribution for the uncertain EMP/CEMP candidates. Here there appears to be a tendency, perhaps statistically significant, for the candidate stars to be more concentrated near the line defining the diagonal cut. We have no ready explanation at present for why the stars in Table 2 would tend to have higher $S$(Ca II K) at a given $S(\mathrm{H}_\epsilon)$.

### 3.2. Application II — A Direct Route to the CEMP Stars

Tables 1 and 2 contain 11 stars with visible CH absorption at SDSS resolution making them candidate CEMP stars. While we deduced the presence of CH by direct inspection of the spectra during the culling phase described in Section 3.1.4, there is obviously a more direct route to CEMP stars via an additional LSP step. We can quickly identify likely C-enhanced EMP stars by taking the 695 stars identified as EMP candidates by the steps in Section 3.1.2 and subjecting them to an additional cut on the depth of the 4308 Å A-X (0,0) band head of $^{12}$CH, $S$(CH 4308). Figure 8 shows the ($S$(Ca II K), $S$(CH 4308)) phase space where we have chosen to make a cut to select only those stars with $S$(CH 4308) $> 0.14$. This cut produces 52 candidate CEMP stars. Once the 52 spectra were culled for defective spectra, spectra whose CH "signal" was merely depression from the far blue wing of $\mathrm{H}_\gamma$, and spectra from stars which otherwise fell outside the bounds of known EMP stars, we were left with 10 of the 11 likely CEMP stars indicated in Tables 1 and 2. The one exception, J080711.80+151458.9, (see Figure 6) was *not* picked up by the cut in ($S$(Ca II K), $S$(CH 4308)) phase space simply because its $S$(CH 4308) value (0.08) was too small to make the cut. We could have made the cut value smaller but that would have been at the expense of having to sort through a significantly larger population of candidate spectra. We will return to this point in Section 4.

We show in Figure 9 the candidate CEMP stars of Table 1 which do not already appear in Figure 6.

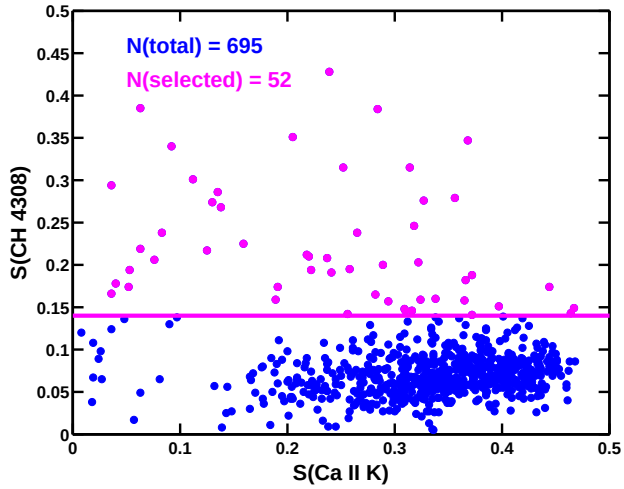


Fig. 8.— Culling potential CEMP candidates. All the stars in the displayed portion of the ($S$(Ca II K),$S$(CH 4308)) plane satisfy the constraints ($1.0 \leq (u\text{-}z) \leq 1.7$), (median S/N) $> 35$), and ($S$(Ca II K) $< 0.6 \times S(\mathrm{H}_\epsilon)$). The stars shown in magenta *additionally* satisfy the constraint that ($S$(CH 4308) $> 0.14$), a cut that extracts high likelihood CEMP candidates. The color-coded "N"s indicate how many stars are in the figure area and how many were sub-selected.

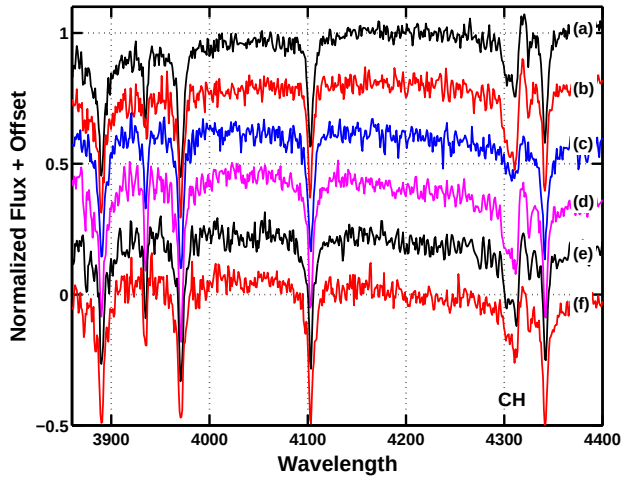


Fig. 9.— The rest of the candidate CEMP stars from Table 1. The spectra have been normalized to their mean flux in the interval [4150,4270]. The stars shown are: (a) J012822.15+225013.4; (b) J081554.26+472947.9; (c) J081754.08+451035.6; (d) J094940.20+270703.6; (e) J125556.07+101715.1; (f) J145409.78-001111.0. The two other CEMP candidates from the table may be found in Figure 6.

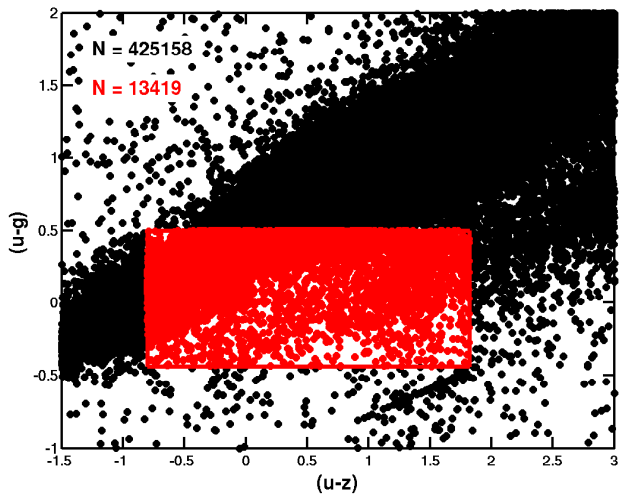


Fig. 10.— First cut in color-color space for CV stars. All the stars present in this portion of phase space are plotted in black. The stars in red were selected using constraints that reflect the distribution of Szkody *et al*'s SDSS CV stars in our downloaded dataset: -0.80 $\leq$ ($u$-$z$) $\leq$ 1.82, and -0.44 $\leq$ ($u$-$g$) $\leq$ 0.50. The color-coded "N"s indicate how many stars are in the figure area and how many were sub-selected.

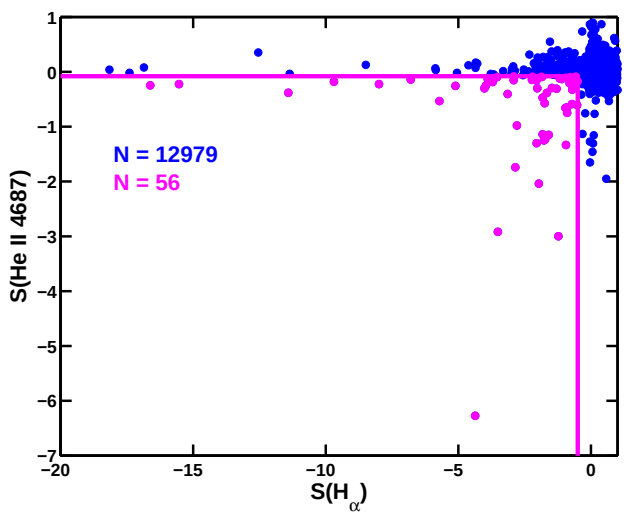


Fig. 11.— Refining color selections with an additional cut in ($S$(H$_\alpha$),$S$(He II 4687)) phase space. We show here primarily the quadrant of this phase space containing stars with both H$_\alpha$ *and* He II 4687 emission ($S$(H$_\alpha$) and $S$(He II 4687) both negative). The stars marked in blue satisfy the cuts we already made in color-color phase spaces: $-0.80 \leq$ ($u$-$z$) $\leq$ 1.82, -0.44 $\leq$ ($u$-$g$) $\leq$ 0.50, and -0.60 $\leq$ ($g$-$r$) $\leq$ 0.90. The stars marked in magenta *additionally* satisfy the constraints ($S$(H$_\alpha$) < -0.5) and $S$(He II 4687) < -0.08). The color-coded "N"s indicate the number of stars selected within the figure bounds.

### 3.3. Application III — Finding CV stars

Emission line spectra are encountered in a variety of SDSS objects: stars, galaxies, and QSOs. In this section we quickly illustrate how the LSP approach can incorporate emission lines by demonstrating a search for cataclysmic variable stars. These objects are conveniently available in our downloaded stellar dataset because they were deliberately targeted by SDSS investigators (Stoughton et al. 2002). To constrain the CV star search area, we first made use of the colors of 117 known SDSS CV stars catalogued by Szkody et al. (2011), made available at the associated WWW site, and present in our downloaded set of SDSS spectra. Figure 10 shows the ($u$-$z$,$u$-$g$) plane where we have selected a subregion characteristic of the 117 known SDSS CV stars. This cut produced 13,419 candidates out of the 425,158 stars in the plane. A very slight additional reduction to 13,014 candidates was accomplished by turning to the ($u$-$z$,$g$-$r$) plane and making another cut informed by the parameters of the Szkody *et al* stars, -0.60 $\leq$ ($g$-$r$) $\leq$ 0.90.

While the two cuts so far have roughly isolated a region occupied by many CV stars in color space, the selected region is also occupied by many other stellar types, especially binaries composed of a white dwarf and an emission line M dwarf. To quickly find stars which were likely to be bona fide CV stars, we adopted the expedient of a selection based on H and He II emission line strengths. While H emission is an expected feature of CV stars, an examination of the Szkody *et al* stars reveals that He II emission is by no means present in all CV stars; see, for example, Szkody et al. (2011, Table 3). However, He II emission is *also* not present in most of the stars contaminating the color selected sample. We therefore sub-selected stars with $S$(He II 4687) $<$ -0.08 and $S(\mathrm{H}_\alpha)$ $<$ -0.5 as a crude vehicle to eliminate most of the contaminating stars. While using He II emission this way certainly must remove some CV stars from our final sample, it does allow us for illustration to more easily get to a robust subset of He II emission CV star candidates. Our selection in the ($S(\mathrm{H}_\alpha)$,$S$(He II 4687)) plane is shown in Figure 11 which shows only stars already selected by the two previous cuts in color spaces.

Fifty-six stars were selected by the last cut and, after culling stars with defective spectra, fully 36 remained as apparently good He II emission CV star candidates. Examination of the SDSS spectral type parameter ("subClass") for these 36 candidates showed that all had been classified "CV" by the SDSS pipeline, a result that is an affirmation of our series of cuts in color and emission line spaces.

While 32 of the 36 candidates are in the Szkody *et al* listing, 4 are not. Of those, 2 have been discussed in the literature as CV stars. The first is J131223.49+173659.2, first noted by Vogel et al. (2008) based on X-ray variability and then confirmed by them based on its light curve and spectra as an AM Her type CV star. The second, J143209.78+191403.4, has only been mentioned in the literature as a CV star by Wei et al. (2013) who recognized it using an interesting data mining approach. They used their Monte Carlo Local Outlier Factor approach to data mine over 656,000 SDSS Data Release 8 stellar spectra in search of spectra that were significant outliers from the norm. This interesting approach, which uses principal component analysis to reduce

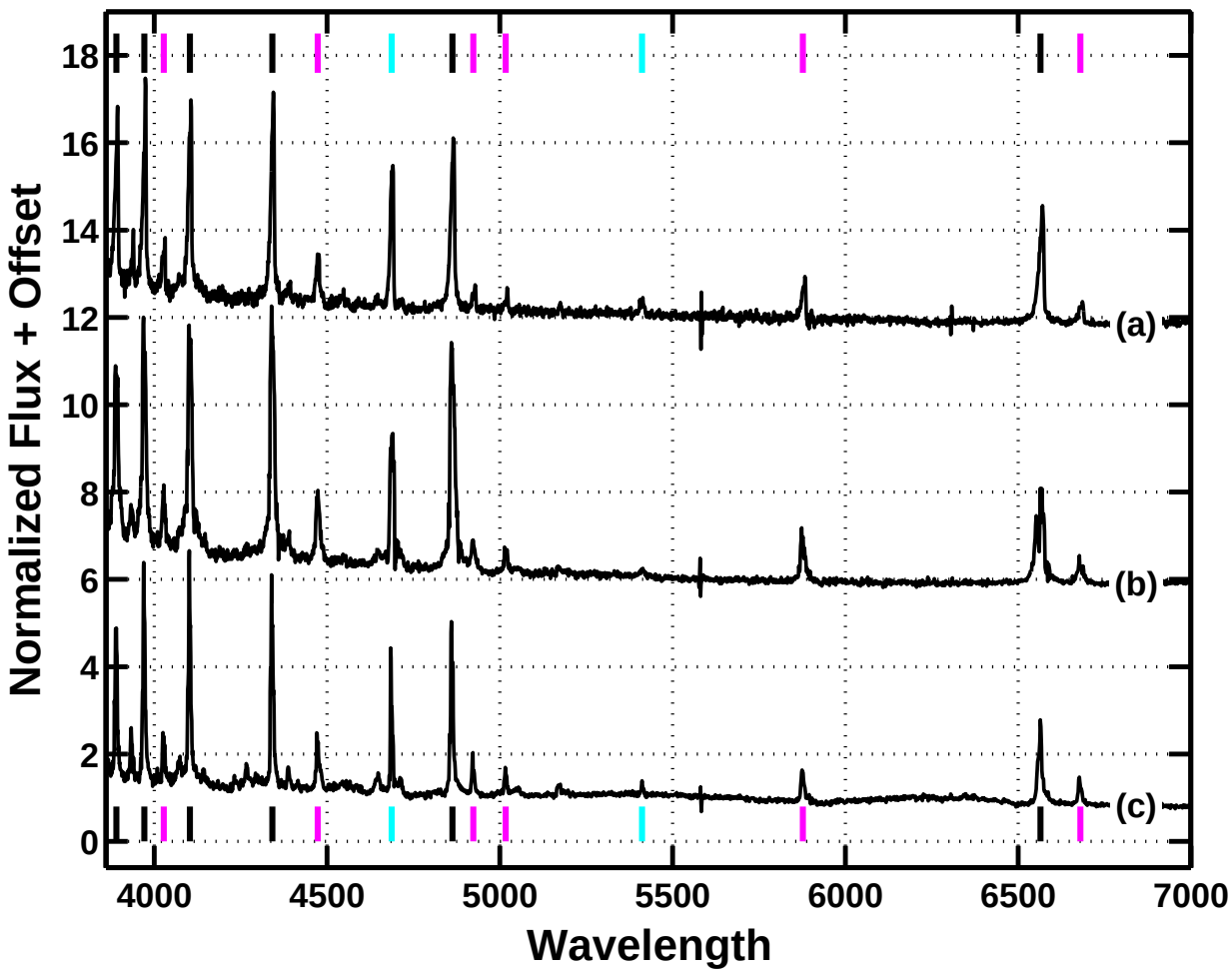


Fig. 12.— Spectra of two "new" candidate CV stars J014227.07+001729.8 (b) and J153023.63+220646.4 (c) compared with the spectrum of the known SDSS CV star J100516.61+694136.5 (a). The wavelengths of the H I Balmer series are marked at the top and bottom by the vertical black bars; the wavelengths of strong He I lines are denoted by magenta bars; the two most prominent lines of He II are marked by cyan bars. Some He II transitions are blended at the resolution of this figure with Balmer lines; for example, the He II 6561.95 is blended with $H_\alpha$. The flux of each star has been normalized to its mean flux in the interval [5250,6500] and plotted with an offset.

each spectrum for processing, produced 5 new CV stars from the DR8 dataset, one of which was J143209.78+191403.4.

It is instructive to understand why our approach did not detect the other four Wei et al. stars. One of these stars, J161935.76+524631.8, did not meet the restrictions of our DR10 download and hence was not considered. Another of the Wei et al. stars, J111715.91+175741.7 did not have valid SDSS photometry and so was not considered by us. Had its $u$-$z$, $u$-$g$ and $g$-$r$ colors been in the selected ranges, it would have been detected by us based on its $S(\mathrm{H}_\alpha)$ and $S$(He II 4687) values. The Wei et al. star J035747.17-063850.8 survived the selection cuts but was eliminated in the culling phase because it did not display convincing evidence of He II 4687 emission when its unfortunately low S/N (median S/N = 5.34) spectrum was closely examined. The fourth star, J140429.36+172359.6, was eliminated by us because its value of $u$-$g$ (-0.467) missed our cut by 0.027 magnitudes. Even had we chosen a more negative value for the lower bound on $u$-$g$, it still would have escaped us because its $S$(He II 4687) measurement (0.041) narrowly fails the cut in that variable. This provides an interesting object lesson since the failure in this instance is due to bad placement of the automatic continuum. Had the continuum been placed "by eye" it would have been lower and the subsequent value of $S$(He II 4687) would have been sufficiently negative ($S$(He II 4687) $\approx 0.3$) to cause the star to be selected. The underlines the importance of good continuum fitting, especially when cuts are made on comparatively small values of $|S(\lambda_i)|$.

The remaining 2 He II emission CV star candidates of the 36 found, J014227.07+001729.8 and J153023.63+220646.4, do not appear to be discussed in the CV star literature according to SIMBAD (Wenger et al. 2000). This may be owing to their rather faint magnitudes. The former star does appear without remarks in the catalog of QSOs complied by Croom et al. (2009) and was presumably recognized by them as stellar. According to the SDSS targeting parameter "sourceType", J014227.07+001729.8 was targeted for SDSS observation as a potential QSO and J153023.63+220646.4 was targeted as a potential white dwarf. Additional details for these two stars appear in Table 3. We show the spectra of these two candidate He II emission CV stars in Figure 12 along with the spectrum of J100516.61+694136.5 which is the CV star in our downloaded Szkody et al reference set which most closely matches the two candidates both in overall spectral distribution and in the relative intensity of individual emission lines. (J100516.61+694136.5, is classified by Wils et al. (2010) as a dwarf nova and a likely magnetic CV star.) Close examination of the three SDSS spectra reveals that they do differ among themselves in the detailed shape and the wavelength shifts of peaks of their strong emission lines.

Table 3. Candidate CV stars

| Star Name | Plate-Fiber-MJD | $g$ | $u$-$g$ | $g$-$r$ | Spectral Type | Notes |
|---|---|---|---|---|---|---|
| J014227.07+001729.8 | 4231-808-55444 | 20.82 | 0.17 | 0.06 | CV | ··· |
| J153023.63+220646.4 | 3949-814-55650 | 19.08 | 0.05 | -0.09 | CV | ··· |

## 4. SHORTCOMINGS AND PLANNED IMPROVEMENTS

A recurrent problem with our approach and, quite likely, any automated approach to data mining spectrum datasets is the necessity to cull through the results "by eye" to eliminate those objects which most definitely do not belong to the desired set of objects but whose spectra and/or other measured parameters lead to their automated capture. The fault may be due to many factors, often unforeseen, including flaws in the spectra themselves and inadequate choices of the constraints used to select the desired objects. The problem can be very formidable. For example, when Rebassa-Mansergas et al. (2010) used a $\chi^2$ approach with templates to cull white dwarf-main sequence binaries from 1.27 million DR6 spectra, they were forced to examine $\approx$ 70,000 spectra "by eye" to remove interloping QSOs, galaxies, etc. With an average of only 5 seconds per spectrum, this is nearly 100 hours of tedious work. Even if the automated approach produces a relatively small number of candidate spectra, the culling of extraneous objects can be onerous if more than a few seconds must be spent per spectrum. As an example, it will be recalled that in our search for CEMP stars (Section 3.2), we set the cutoff to be $S(\text{CH } 4308) > 0.14$. Had we lowered the cutoff to $S(\text{CH } 4308) > 0.07$, we would have captured at least the one CEMP we had missed from Table 1 and probably other similar stars. However, this choice would have produced 356 stars to be individually examined, nearly 7 $\times$ the previous number. Since very much more than 5 seconds per spectrum was always required to decide whether or not a star was an EMP candidate, the additional effort would not have been trivial.

The user is always faced with making a compromise: finding more candidates for the desired objects versus more time spent examining individual spectra in order to remove false candidates. We believe that our approach, coupling LSPs with the hyperwall, with its *immediate* visual feedback regarding exactly what is being selected by the one's choices gives users a significant advantage. If one finds that too many undesirable objects are being selected, hopefully one can devise constraints that will eliminate the bulk of these.

In extreme cases, being able to devise better constraints may require changes to the spectrum processing pipeline and the introduction of new spectrum measures into the mix. For example, we had adopted linear interpolation to handle gaps in the SDSS spectra, primarily as a tool to simplify continuum fitting. This led to false measurements of $S(\lambda_i)$ and $D(\lambda_i)$ whenever line features landed in those gaps. Such erroneous measures were responsible for a sizable fraction of the false EMP candidates that had to be culled by eye. Our future pipelines will record "no-measurement" when feature wavelengths fall within a spectrum gap.

Another improvement that is useful and easy to implement concerns the wavelengths of molecular band heads. The current set of band head wavelengths, drawn as it is from the molecular spectroscopy literature, is strictly appropriate only to high resolution spectra. When the same band systems are viewed at much lower resolutions like those of the SDSS spectra, the apparent position of maximum band head absorption shifts, often by a quite noticeable amount. The effect can become even more complex because it is also partially dependent on the strength of the molec-

ular feature which usually varies greatly with spectral type. It appears that a better approach than simply using the band head wavelengths might be to pick by visual examination non-band-head wavelengths comfortably offset from the band head as it degrades. While $S(\lambda_i)$ at these offset positions might not be as large valued as those at the band heads, the overall effect with regard to monitoring molecular line strength is probably positive. Selection of the offset wavelengths could be accomplished with comparatively little effort using appropriate SDSS spectra as a guide.

Finally, it became apparent when sorting through the initial candidates in Sections 3.1.3, 3.1.4, and 3.2 that measures of strong line width could be very helpful in the culling process. Such measures would have greatly simplified the elimination of hotter stars from the candidate EMP/CEMP lists and helped excise stars whose $H_\gamma$ wing was sufficient to produce a false $S$(CH 4308) signal. Since line width measures could easily have other applications as well, it is our goal to add a simple width measure for strong lines to the next pipeline version.

## 5. SUMMARY

We have described a new approach to the problem of exploring the contents of very large datasets of spectra. We showed how the pairing of linked scatter plots with the visual display capabilities of NASA's hyperwall creates a powerful and highly flexible spectrum data exploration tool. To exploit this tool, we devised a spectrum reduction pipeline to produce measurements ($S$ and $D$) of atomic and molecular feature strengths in each individual spectrum of a spectrum archive. This pipeline was used to process 569,738 stellar spectra downloaded from the SDSS DR10 archive, producing nearly 2 billion individual feature strength measurements to explore with LSPs on the hyperwall. The hyperwall allows us to construct and examine multiple phase spaces from these feature strength measurements and other parameters. The LSP capability on the hyperwall then permits complex data exploration by allowing the user to subselect objects with multiple desired characteristics using straightforward graphical selection tools provided on each hyperwall display.

Using simple selections in just two phase spaces,($u$-$z$,median S/N) and ($S(H_\epsilon)$,$S$(Ca II K)), we have revealed 57 new likely EMP/CEMP candidates (Table 1) and 13 uncertain EMP/CEMP candidates (Table 2). We show how one can go directly to revealing candidate CEMP stars without visually inspecting spectra simply by adding one more selection in the additional ($S$(Ca II K), $S$(CH 4308)) phase space. Finally, we demonstrate how one can reveal CV stars candidates with He II emission lines using simple selections in only three phase spaces,($u$-$z$,$u$-$g$), ($u$-$z$,$g$-$r$), and ($S(H_\alpha)$,$S$(He II 4687)).

The above examples are simple applications of our LSP/hyperwall tool chosen to illustrate the general principles while using only a small number of phase spaces. More complex selection choices involving a larger number of phase spaces are trivially accomplished. We believe that our approach is widely applicable to large spectrum data archives of stellar, galactic, and QSO spectra, possibly

even planetary spectra, and we are interested in hearing from researchers who may have suitable search problems.

The authors wish to sincerely thank Ruth Peterson, David Schwenke, and Karen Huyser for their insightful comments on the draft of this paper. DFC is indebted to Richard O. Gray for his help in interpreting the WD/dMe spectra encountered in the search for CV stars. Funding for SDSS-III has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, and the U.S. Department of Energy Office of Science. The SDSS-III web site is http://www.sdss3.org/. SDSS-III is managed by the Astrophysical Research Consortium for the Participating Institutions of the SDSS-III Collaboration including the University of Arizona, the Brazilian Participation Group, Brookhaven National Laboratory, Carnegie Mellon University, University of Florida, the French Participation Group, the German Participation Group, Harvard University, the Instituto de Astrofisica de Canarias, the Michigan State/Notre Dame/JINA Participation Group, Johns Hopkins University, Lawrence Berkeley National Laboratory, Max Planck Institute for Astrophysics, Max Planck Institute for Extraterrestrial Physics, New Mexico State University, New York University, Ohio State University, Pennsylvania State University, University of Portsmouth, Princeton University, the Spanish Participation Group, University of Tokyo, University of Utah, Vanderbilt University, University of Virginia, University of Washington, and Yale University. This research has made use of the SIMBAD database, operated at CDS, Strasbourg, France. MATLAB© 2015 The MathWorks, Inc. MATLAB and Simulink are registered trademarks of The MathWorks, Inc. See www.mathworks.com/trademarks for a list of additional trademarks. Other product or brand names may be trademarks or registered trademarks of their respective holders.

*Facilities:* Sloan

## REFERENCES

Ahn, C. P., Alexandroff, R., Allende Prieto, C., et al. 2014, ApJS, 211, 17

Allende Prieto, C., Beers, T. C., Wilhelm, R., et al. 2006, ApJ, 636, 804

Aoki, W., Beers, T. C., Sivarani, T., et al. 2008, ApJ, 678, 1351

Aoki, W., Beers, T. C., Lee, Y. S., et al. 2013, AJ, 145, 13

Bailer-Jones, C. A. L. 2000, A&A, 357, 197

Beers, T. C. 2014, Mem. Soc. Astron. Italiana, 85, 525

Beers, T. C., & Carollo, D. 2008, in American Institute of Physics Conference Series, Vol. 990, First Stars III, ed. B. W. O'Shea & A. Heger, 104–108

Beers, T. C., & Carollo, D. 2012, in European Physical Journal Web of Conferences, Vol. 19, European Physical Journal Web of Conferences, 2001

Beers, T. C., & Christlieb, N. 2005, ARA&A, 43, 531

Behara, N. T., Bonifacio, P., Ludwig, H.-G., et al. 2010, A&A, 513, A72

Bijaoui, A., Recio-Blanco, A., de Laverny, P., & Ordenovic, C. 2012, Statistical Methodology, 9, 55

Boeche, C., Siebert, A., Williams, M., et al. 2011, AJ, 142, 193

Bonifacio, P., Sbordone, L., Caffau, E., et al. 2012, A&A, 542, A87

Bonifacio, P., Caffau, E., François, P., et al. 2011, Astronomische Nachrichten, 332, 251

Bonifacio, P., Caffau, E., Spite, M., et al. 2015, A&A, 579, A28

Caffau, E., Bonifacio, P., François, P., et al. 2011a, Nature, 477, 67

—. 2011b, A&A, 534, A4

—. 2012, A&A, 542, A51

—. 2013, A&A, 560, A15

Caffau, E., Sbordone, L., Bonifacio, P., et al. 2014, Mem. Soc. Astron. Italiana, 85, 222

Cannon, A. J., & Pickering, E. C. 1918, Annals of Harvard College Observatory, 91, 1

Carollo, D., Beers, T. C., Bovy, J., et al. 2012, ApJ, 744, 195

Chiappini, C., Minchev, I., Anders, F., et al. 2015, Astrophysics and Space Science Proceedings, 39, 111

Cohen, M. 1980, ApJ, 238, L81

Cooke, R. J., & Madau, P. 2014, ApJ, 791, 116

Covey, K. R., Ivezić, Ž., Schlegel, D., et al. 2007, AJ, 134, 2398

Croom, S. M., Richards, G. T., Shanks, T., et al. 2009, MNRAS, 392, 19

Dierickx, M., Klement, R., Rix, H.-W., & Liu, C. 2010, ApJ, 725, L186

Frebel, A., Chiti, A., Ji, A. P., Jacobson, H. R., & Placco, V. M. 2015, ApJ, 810, L27

Frebel, A., & Norris, J. E. 2015, ARA&A, 53, 631

Gänsicke, B. T., Koester, D., Marsh, T. R., Rebassa-Mansergas, A., & Southworth, J. 2008, MNRAS, 391, L103

Gazis, P. R., Levit, C., & Way, M. J. 2010, PASP, 122, 1518

Gray, R. O., & Corbally, J., C. 2009, Stellar Spectral Classification (Princeton University Press)

Henze, C. 1998, in Proceedings of the Conference on Visualization '98, VIS '98 (Los Alamitos, CA, USA: IEEE Computer Society Press), 87–94

Idehara, H., & Carbon, D. F. 2004, in Bulletin of the American Astronomical Society, Vol. 36, American Astronomical Society Meeting Abstracts, 1602

Kordopatis, G., Wyse, R. F. G., & Binney, J. 2016, ArXiv e-prints, arXiv:1601.02673

Kordopatis, G., Gilmore, G., Steinmetz, M., et al. 2013, AJ, 146, 134

Kramida, A., Ralchenko, Yu, Reader, J., & and NIST ASD Team (2013). 2014, NIST Atomic Spectra Database (ver. 5.1), [Online]. Available: `http://physics.nist.gov/asd` [2014, April 15]. National Institute of Standards and Technology, Gaithersburg, MD., ,

Lai, D. K., Rockosi, C. M., Bolte, M., et al. 2009, ApJ, 697, L63

Lee, Y. S., Beers, T. C., Sivarani, T., et al. 2008, AJ, 136, 2022

Lee, Y. S., Beers, T. C., Masseron, T., et al. 2013, AJ, 146, 132

Li, X., Zhao, G., Chen, Y.-Q., & Li, H.-N. 2014, Research in Astronomy and Astrophysics, 14, 1423

Manteiga, M., Ordóñez, D., Dafonte, C., & Arcay, B. 2010, PASP, 122, 608

Martell, S. L., & Shetrone, M. D. 2013, MNRAS, 430, 611

MATLAB. 2011, version 7.13.0.564 (R2011b) (Natick, Massachusetts: The MathWorks Inc.)

Merrill, P. W. 1947, ApJ, 106, 274

Perrin, G., Coudé du Foresto, V., Ridgway, S. T., et al. 1998, in Astronomical Society of the Pacific Conference Series, Vol. 154, Cool Stars, Stellar Systems, and the Sun, ed. R. A. Donahue & J. A. Bookbinder, 1636–1643

Peters, C. S., & Thorstensen, J. R. 2005, PASP, 117, 1386

Placco, V. M., Frebel, A., Lee, Y. S., et al. 2015, ApJ, 809, 136

Re Fiorentin, P., Bailer-Jones, C. A. L., Lee, Y. S., et al. 2007, A&A, 467, 1373

Rebassa-Mansergas, A., Gänsicke, B. T., Schreiber, M. R., Koester, D., & Rodríguez-Gil, P. 2010, MNRAS, 402, 620

Recio-Blanco, A., Bijaoui, A., & de Laverny, P. 2006, MNRAS, 370, 141

Recio-Blanco, A., de Laverny, P., Allende Prieto, C., et al. 2016, A&A, 585, A93

Sánchez-Blázquez, P., Peletier, R. F., Jiménez-Vicente, J., et al. 2006, MNRAS, 371, 703

Sandstrom, T. A., Henze, C., & Levit, C. 2003, in International Conference on Coordinated and Multiple Views in Exploratory Visualization, 2003. Proceedings

Sanford, R. F. 1949, PASP, 61, 261

Sbordone, L., Bonifacio, P., Caffau, E., & Ludwig, H.-G. 2012, in Astronomical Society of the Pacific Conference Series, Vol. 458, Galactic Archaeology: Near-Field Cosmology and the Formation of the Milky Way, ed. W. Aoki, M. Ishigaki, T. Suda, T. Tsujimoto, & N. Arimoto, 69–72

Stoughton, C., Lupton, R. H., Bernardi, M., et al. 2002, AJ, 123, 485

Szkody, P., Anderson, S. F., Brooks, K., et al. 2011, AJ, 142, 181

Theodossiou, E., & Danezis, E. 1991, Ap&SS, 183, 91

Vogel, J., Byckling, K., Schwope, A., et al. 2008, A&A, 485, 787

Wei, P., Luo, A., Li, Y., et al. 2013, MNRAS, 431, 1800

Wenger, M., Ochsenbein, F., Egret, D., et al. 2000, A&AS, 143, 9

Wilhelm, R., Beers, T. C., & Gray, R. O. 1999, AJ, 117, 2308

Wils, P., Gänsicke, B. T., Drake, A. J., & Southworth, J. 2010, MNRAS, 402, 436

Wilson, D. J., Gänsicke, B. T., Koester, D., et al. 2014, MNRAS, 445, 1878

Wyse, R. F. G. 2016, ArXiv e-prints, arXiv:1604.04745

Xing, Q. F., & Zhao, G. 2014, ApJ, 790, 33

---

This preprint was prepared with the AAS LaTeX macros v5.2.